
Subject: Merging rounds of survey

Posted by [nogtd](#) on Mon, 27 Jan 2014 13:18:39 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear all,

I have merged the 4 available datasets for Burkina Faso in one single dataset (on Stata) to make some repeated cross-sectional analysis. However, I have noticed that the regions are not coded the same across the rounds of surveys (for instance, when I cross-tabulate the round and the region variable, I have clearly a discrepancy, or a displacement of some regions codes from one round to another). Besides, for the last 2 rounds, the values for the regions are not labelled any more.

I have been told it was quite usual but I cannot find on the DHS website, the DHS recode variables manual, or any forum topics how to recode them. I would need the regions to make the rounds coincide.

Does anyone have a clue how to correct this ? Thank you in advance, and sorry for my English !

kind regards,

N

Subject: Re: Merging rounds of survey

Posted by [user-rhs](#) on Wed, 29 Jan 2014 22:44:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi N,

I have run into this problem before with housing materials (the vbls I looked at) when I tried to stack data from multiple waves into a single dataset. As with your datasets, the vbls were not coded consistently across survey years.

The way I would go about this is to create a variable that standardises the coding of the regions for each wave before doing the merge. Check with the codebook (the FRQ or FRW file that comes in the zip file w/ the dataset) for coding.

For example, for Bangladesh 2004, wall material is coded as follows:

```
. label list v128
```

```
v128:
```

```
10 natural
```

```
11 jute/bamboo/mud (katcha)
```

```
20 rudimentary
```

```
21 wood
```

```
30 finished
```

```
31 brick/cement
```

```
32 tin
```

96 other

In Bangladesh 2007, same vbl is coded as:

. label list v128

v128:

- 11 no walls
- 12 cane / palm / trunks
- 13 dirt
- 22 bamboo with mud
- 23 stone with mud
- 24 plywood
- 25 cardboard
- 31 tin
- 32 cement
- 33 stone with lime / cement
- 34 bricks
- 35 wood planks / shingles
- 96 other
- 97 non de jure resident

In each of the datasets (2004 and 2007), I would create a new variable called "wall" (or v128_2, or whatever you want) from v128 coded as:

- 1 - no walls
- 2 - Jute/bamboo
- ...
- 7 - Tin
- etc.

Of course, you may have to make some executive decisions (i.e., is jute/bamboo/mud equivalent to bamboo with mud?), but administrative regions, particularly if they're at higher levels (e.g. provinces), are probably pretty stable over time. If the administrative regions have changed over time (e.g. a large province is divided into 2 provinces), I personally would recode the relatively new provinces into the old province's name.

HTH,
rhs
